JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

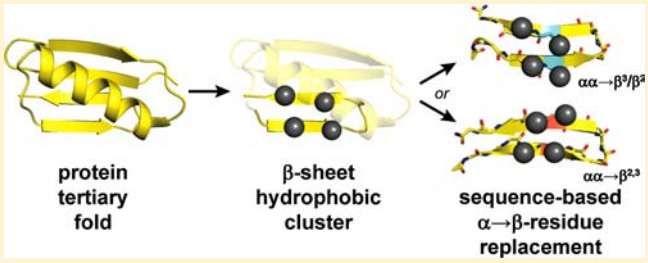
Design Strategies for the Sequence-Based Mimicry of Side-Chain Display in Protein β-Sheets by α/β-Peptides

George A. Lengyel and W. Seth Horne*

Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States

Supporting Information

ABSTRACT: The sophistication of folding patterns and functions displayed by unnatural-backbone oligomers has increased tremendously in recent years. Design strategies for the mimicry of tertiary structures seem within reach; however, a general method for the mimicry of sheet segments in the context of a folded protein is an unmet need preventing realization of this goal. Previous work has shown that 1→1 $\alpha\rightarrow\beta$-residue substitutions at cross-strand positions in a hairpin-forming α-peptide sequence can generate an α/β-peptide analogue that folds in aqueous conditions but with a change in side-chain display relative to the natural sequence; this change would prevent application of single β-residue substitutions in a larger protein. Here, we evaluate four different substitution strategies based on replacement of $\alpha\alpha$ dipeptide segments for the ability to retain both sheet folding encoded by a parent α-peptide sequence as well as nativelike side-chain display in the vicinity of the β-residue insertion point. High-resolution structure determination and thermodynamic analysis of folding by multidimensional NMR suggest that three of the four designs examined are applicable to larger proteins.



INTRODUCTION

Unnatural-backbone oligomers that fold like proteins, often termed "foldamers,"[1] offer the exciting possibility of manifesting proteinlike biological functions on protease-resistant scaffolds.[2] Early work on the use of these molecules to mimic natural proteins relied heavily on structure-based approaches.[2f] Such methods depend on the synthesis and characterization of diverse, de novo designed sequences; as some understanding of the fundamental relationships between sequence and folding in a given unnatural backbone emerges, rational design can be used to create an entity with both a desirable biological function and a protease-resistant backbone. Recent years have seen the emergence of an alternate approach for the design of functional foldamers in which a natural protein sequence serves as the starting point for design of its own mimic. Termed "sequence-based design," the latter method is best exemplified in its application to prepare α/β-peptides[3] (oligomers composed of mixtures of natural α-amino acid residues and unnatural β-amino acid residues) that fold and function like proteins.

The basic idea underlying sequence-based design is that natural proteins will manifest their inherent folding behavior when a native sequence of side chains is displayed on an unnatural backbone, provided the backbone modification is made judiciously. The first applications of sequence-based design focused on α-helical prototype sequences and led to significant advances in the development of oligomers with complex folding patterns and functions, including formation of helix-bundle quaternary structures[4] and high-affinity binding to biological receptors involved in HIV-cell fusion[5] and apoptosis.[6] More recently, sequence-based design has begun to expand beyond helix mimicry to other structural contexts, such as sheet[7] and a fold with less regular secondary structure.[8]

The development of unnatural oligomers that mimic the complex tertiary folds of proteins is a frontier scientific challenge. We believe that sequence-based modification of natural proteins can provide a general and readily implemented strategy to accomplish this goal. Success in such an approach requires rules for the modification of all the secondary structure elements commonly encountered in proteins: helices, tight turns, loops, and sheets. Design principles for helix mimicry by $\alpha\rightarrow\beta$ residue substitution are now well-established,[4−6] and other methods for helix replacement have shown promise.[9] A number of turn replacements have been reported to be effective in folded proteins,[10] and loops in a tertiary structure can be tolerant to significant modification.[11] Building on pioneering early work on sheet folding behavior of β-residues in organic solvent,[12] we have recently begun to examine the applicability of sequence-based $\alpha\rightarrow\beta$ backbone modification to the mimicry of protein sheets in water (Figure 1).[7]

In previous work, we compared the sheet-folding propensities of 16 different β-residues when inserted into a short α-peptide hairpin model system; the β-residues had systematically varied substitution pattern, stereochemistry, and ring constraint.[7] Of the monomers examined, only three led to hairpin folding in aqueous buffer when introduced as single $\alpha\rightarrow\beta$ replacement at cross-strand positions the prototype α-peptide.

Received: June 28, 2012
Published: September 4, 2012

   dx.doi.org/10.1021/ja306311r | J. Am. Chem. Soc. 2012, 134, 15906−15913

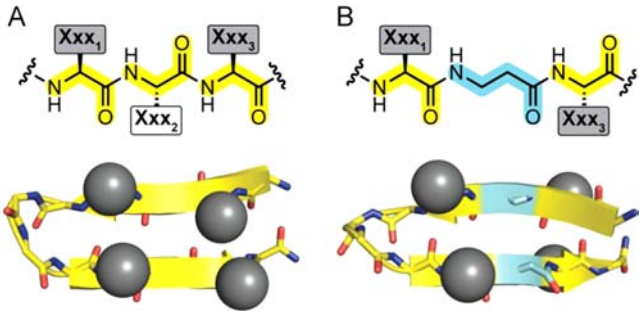


Figure 1. Change in side-chain display resulting from 1→1 $\alpha\rightarrow\beta$ residue substitution in each strand of a β-hairpin. Residues i and $i + 2$ (gray boxes and spheres) are on the same face in the natural backbone (A) but on opposite faces in the unnatural backbone (B). Coordinates are from published NMR solution structures of an α-peptide hairpin and an analogue resulting from two $\alpha\rightarrow\beta$ residue substitutions[7].

Unfortunately, the 1→1 $\alpha\rightarrow\beta$ residue substitution was found unsuitable for modification of a larger protein sheet, due to a fundamental change in side-chain display resulting from the backbone modification (Figure 1).

The hydrogen-bonding pattern of a β-residue in an extended strand typically differs from that of an α-residue in the same context.[7,12] As a result, the side chains of α-residues flanking the point of $\alpha\rightarrow\beta$ substitution end up on opposite faces of the strand rather than residing on the same face as they do in the natural backbone (Figure 1). None of the 16 monomers examined in our prior study were able to suppress this change in side-chain display resulting from single $\alpha\rightarrow\beta$ replacement. Although tolerated in a small hairpin, this change would abolish folding in the context of a larger tertiary structure as it would fundamentally alter the side-chain composition of the hydrophobic core of the protein. We therefore sought to devise and test alternate β-residue substitution strategies that would retain both sheet folding of a parent α-peptide sequence as well as nativelike side-chain display in the vicinity of the insertion point of the unnatural residue. We report here several designs that accomplish both these goals.

MATERIALS AND METHODS

Monomer Synthesis. For full experimental details and characterization data, see the Supporting Information. β^3-Monomers were purchased as the free amino acid and Fmoc protected. β^2-Amino acids were synthesized using a route based on a diastereoselective Mannich reaction as the key step.[13] $\beta^{2,3}$-Amino acid monomers were prepared by either ring-opening of enantiomerically enriched thiazinones[14] or diastereoselective conjugate addition to an α,β-unsaturated ester.[15]

Peptide Synthesis. Peptides were synthesized using microwave-assisted Fmoc solid-phase synthesis techniques on a MARS microwave reactor (CEM). NovaPEG Rink amide resin or H-Glu(*t*Bu) HMPB NovaPEG resin was used as the solid support. Couplings were carried out in NMP with a 2 min ramp to 70 °C and a 4 min hold at that temperature, using 4 equiv of Fmoc-protected amino acid, 4 equiv of HCTU, and 6 equiv of DIEA. Deprotections were performed with a 2 min ramp to 80 °C followed by a 2 min hold at that temperature, using an excess of 20% 4-methylpiperidine in DMF. After each coupling or deprotection cycle, the resin was washed three times with DMF. Double coupling was performed at sequence positions following proline or β-residues. N-terminal acetylation, when present, was carried out on resin by treatment with 8:2:1 v/v/v DMF/DIEA/Ac_2O.

Prior to cleavage from the resin, peptides were washed three times each with DMF, dichloromethane, and methanol, and then dried. Peptide cleavage was performed using 95% trifluoroacetic acid (TFA), 2.5% triisopropylsilane, and 2.5% water. Cysteine-containing peptides were purified, lyophilized, dissolved in 10 mM phosphate buffer (pH 8.9, 5% v/v DMSO), stirred until analytical high-performance liquid chromatography (HPLC) and mass spectrometry (MS) showed complete conversion to the cyclic disulfide (1−2 d), and then repurified. Purification was accomplished by preparative HPLC on a C_{18} column using gradients between 0.1% TFA in water and 0.1% TFA in acetonitrile. All peptides were >95% pure as determined by analytical HPLC on a C_{18} column. Identities were confirmed by mass spectrometry using a Voyager DE Pro matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) instrument (Supporting Information, Table S1).

NMR Sample Preparation and Data Collection. NMR samples were prepared by dissolving peptide in 750−850 μL of degassed 50 mM phosphate, 9:1 H_2O/D_2O, pH 6.3 (uncorrected for the presence of D_2O) to a final concentration of 0.2−1 mM. 3-(Trimethylsilyl)-1-propanesulfonic acid sodium salt (DSS, 50 mM in water) was added to a final concentration of 0.2 mM. Each solution was passed through a 0.2 μm syringe filter, transferred to an NMR tube, and stored until analysis. The NMR tube headspace was purged with a stream of nitrogen prior to capping.

NMR spectra of peptides were recorded on a Bruker Avance-700 spectrometer. Chemical shifts are reported relative to DSS (0 ppm). Total correlation spectroscopy (TOCSY), nuclear Overhauser enhancement spectroscopy (NOESY), and correlation spectroscopy (COSY) pulse programs used excitation-sculpted gradient-pulse solvent suppression. All experiments were obtained using 2048 data points in the direct dimension and 512 data points in the indirect dimension. TOCSY spectra were acquired with a mixing time of 80 ms and NOESY spectra were acquired with a mixing time of 200 ms. NMR measurements were performed at a temperature of 293 K unless otherwise noted. We measured linear hairpin peptides with unnatural backbones (**6a**−**9a**) at 278 K to maximize folded population and facilitate comparison of folded stability. Natural backbone peptide **5a** was measured at both 278 and 293 K. NMR data at 293 K were used for comparison among the α-peptide series (**1a**−**5a**), while data at 278 K were used for comparison with the unnatural backbone series (**5a**−**9a**).

NMR Data Analysis and Structure Determination. The Sparky software package (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) was used to analyze two-dimensional (2D) NMR data. Backbone chemical shift assignments were generated (Supporting Information, Tables S2−S9), and each cyclic peptide was analyzed for qualitative nuclear Overhauser effects (NOEs) indicative of folding. Peptides that showed a high degree of folding were fully assigned (Supporting Information, Tables S10−S13) and inter-residue NOEs were tabulated. NOE integration values were converted to distance restraints using eq 1:

$$I = cr^6 \quad (1)$$

where I is NOE intensity, c is a constant (determined using resolved diastereotopic CH_2 groups from Tyr, Gly, and/or Phe), and r is distance.[16] The distances were then sorted and classified as strong (≤2.7 Å), medium (≤3.5 Å), weak (≤4.5 Å), or very weak (≤5.5 Å) to generate distance restraints used for structure determination (Supporting Information, Tables S14−S17).

The Crystallography and NMR system (CNS) software package was used to generate three-dimensional (3D) structures.[17] Patches were written for β-residues and accompanying linkages. Distance restraints calculated as above were used in 100 simulated annealing runs using default CNS parameters for protein NMR. Structures including NOE distance-restraint violations >0.5 Å were discarded, and the 10 lowest energy structures were obtained. The minimum energy average of these 10 structures was inspected to identify H-bonding contacts. These contacts were then included in an additional restraint file, and the annealing process was repeated over 500 runs to generate an ensemble of 10 low-energy structures and a minimized average structure for each peptide. In the case of peptides where the NMR ensemble had more than one distinct family of structures, minimized average coordinates for each family were calculated.

Calculation of Folding Equilbria by NMR. Fraction folded from chemical shift deviation ($f_{H\alpha}$) was calculated using experimentally determined H_α chemical shifts ($\delta_{H\alpha}$) in eq 2:

$$f_{H\alpha} = \frac{\delta_{H\alpha,observed} - \delta_{H\alpha,unfolded}}{\delta_{H\alpha,folded} - \delta_{H\alpha,unfolded}} \quad (2)$$

where $\delta_{H\alpha,observed}$ is the chemical shift of a particular H_α in the unknown peptide (e.g., **1a**), $\delta_{H\alpha,unfolded}$ its chemical shift in an N- or C-terminal fragment (e.g., **1c** or **1d**), and $\delta_{H\alpha,folded}$ its chemical shift in a disulfide-bridged cyclic analogue (e.g., **1b**). Reported values of f are averages calculated using chemical shift data for residues 4, 11, and 13.

Fraction folded from separation of diasterotopic Gly H_α's (f_{Gly}) was calculated using eq 3:

$$f_{Gly} = \frac{\Delta\delta_{H\alpha/H\alpha',observed}}{\Delta\delta_{H\alpha/H\alpha',folded}} \quad (3)$$

where $\Delta\delta_{H\alpha/H\alpha',observed}$ is the chemical shift difference between Gly H_α's in an unknown peptide (e.g., **1a**) and $\Delta\delta_{H\alpha/H\alpha',folded}$ the corresponding difference in a disulfide-bridged cyclic analogue (e.g., **1b**).

The equilibrium constant for the folding equilibrium (K_{fold}) and corresponding free energy of folding ($\Delta G°_{fold}$) were calculated from fraction folded using eqs 4 and 5:

$$K_{fold} = \frac{f}{1-f} \quad (4)$$

$$\Delta G_{fold} = -RT\ln(K_{fold}) \quad (5)$$

Experimental uncertainty for a folded population determined by H_α chemical shift deviation ($f_{H\alpha}$) was estimated using the standard deviation of the mean for populations based on residues 4, 11, and 13. Error for a folded population determined by Gly H_α separation (f_{Gly}) was estimated by assuming 0.01 ppm error in NMR peak assignment. The above values were used in standard error propagation based on eqs 3−5 to give uncertainties for f_{Gly}, K_{fold}, and ΔG_{fold}. A lower bound for the folded population of peptide **8a** was calculated using eq 3, an estimated minimum measurable glycine separation value of 0.03 ppm, and an estimated value for the fully folded state of 0.322 ppm (average observed for peptides **5b**, **6b**, **7b**, and **9b**).

RESULTS AND DISCUSSION

Backbone-Substitution Strategies for Mimicry of Side-Chain Display in a Protein Sheet. As detailed in our prior work, single $\alpha\rightarrow\beta$ residue replacement at matching positions along each strand of a hairpin-forming α-peptide can lead to an analogue that retains folding in buffered aqueous solution, albeit with a change in side-chain display (vide supra).[7] Of 16 β-residue types examined in that study, the two with the highest propensity for manifesting the sheet fold of the parent α-peptide sequence were a singly substituted β^3-residue and a doubly substituted *syn*-$\beta^{2,3}$-residue. Based on the above observations, we conceived four design strategies (Figure 2) that we envisioned would lead to nativelike display of side chains after sequence-based modification of a larger α-peptide sheet.

By contrast to the 1→1 ($\alpha\rightarrow\beta$) residue substitution examined previously, designs I−IV represent four different ways of replacing an $\alpha\alpha$ dipeptide with one or more β-residues. The four designs are meant to explore the impact of two orthogonal variables on sheet folding after modification: (1) backbone length relative to the replaced $\alpha\alpha$ segment and (2) backbone conformational preorganization in the introduced β residue(s). The variable of backbone length is addressed by applying 2→1 ($\alpha\alpha\rightarrow\beta$) or 2→2 ($\alpha\alpha\rightarrow\beta\beta$) residue substitution. The role of conformational preorganization is examined by incorporating flexible β^3- or β^2-residues versus more constrained *syn*-$\beta^{2,3}$-monomers.

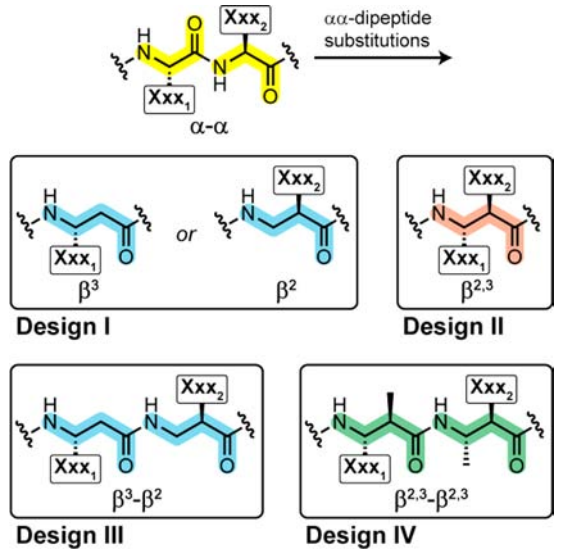


Figure 2. Four substitution strategies for the replacement of $\alpha\alpha$ dipeptide segments in a β-strand. Designs I and II involve 2→1 substitution ($\alpha\alpha\rightarrow\beta$), while III and IV involve 2→2 substitution ($\alpha\alpha\rightarrow\beta\beta$). Designs II and IV employ conformationally constrained *syn*-$\beta^{2,3}$-monomers, while I and III use more flexible β^3- and/or β^2-residues.

In design I, an $\alpha\alpha$ dipeptide is replaced by a single β-residue (β^2 or β^3), resulting in the deletion of two atoms from the protein backbone (amide N and carbonyl C). One side chain is also omitted, and the retained side chain (i.e., Xxx_1 or Xxx_2) at a given site is determined by which residue from the replaced dipeptide is more intimately tied to folding. In design II, two α-residues are condensed into a single conformationally restrained *syn*-$\beta^{2,3}$-monomer, and both parent side chains are retained. In designs III and IV, an $\alpha\alpha$ dipeptide is replaced by a two-residue $\beta\beta$ segment, resulting in the insertion of two additional backbone C atoms. The positioning of side chains is selected to retain the up/back pattern along the strand. Design III accomplishes this through a $\beta^3\beta^2$-segment, while design IV makes use of $\beta^{2,3}$-residues with extra structure-promoting methyl groups on spacer C atoms. In all cases, the absolute stereochemistry of the β-residue(s) employed was selected to best match the L configuration of the replaced $\alpha\alpha$ dipeptide.

Model System Selection and Design of α/β-Peptide Variants. With hypotheses about four possible strategies for sheet $\alpha\rightarrow\beta$ backbone modification in hand, we next looked for a model system[18] in which to systematically compare designs I−IV. We had several requirements for properties of an ideal model sequence. First, we sought an oligomer small enough to be amenable to analysis by ^{1}H-detected multidimensional NMR of samples lacking isotopic labels. Second, we needed a sheet-forming sequence with appropriate folded stability. Our previous work provided some qualitative insights into relative folding propensities of different types of β-residues incorporated into an α-peptide hairpin with an unnatural turn-promoting D-Pro;[7] however, rigorous thermodynamic analysis of folding free energies was precluded there by lack of proper controls to quantify folded population. We sought a system in the present study that would provide a folded hairpin population of 60−70% in the prototype sequence with ready access to related control sequences for 0% and 100% folded

dx.doi.org/10.1021/ja306311r | J. Am. Chem. Soc. 2012, 134, 15906−15913

states. This range would allow for measurable population changes by NMR due to either increase or decrease in folded stability resulting from backbone substitution.

Peptide **1a** (Figure 3), the 16 residue C-terminal segment of the first immunoglobulin-binding domain of Streptococcal protein G (GB1), has been shown to form a β-hairpin in aqueous solution;[19] however, it posed some problems that prevented its use in the present study. The folded state of peptide **1a** is only marginally stable, with a reported hairpin population of ~40% at 278 K.[19] Moreover, in our hands, the large number of threonine residues led to poor chemical-shift dispersion and complicated NMR analysis of its folding equilibrium. In an effort to find a model sequence with more favorable properties, we examined peptides **2a**–**5a** (Figure 3), a series of mutants of sequence **1a**. Peptides **2a** and **4a** are previously reported variants of **1a** with improved folded stability.[20] Sequence **3a** is a mutant of **2a** with three modifications intended to destabilize the folded state: replacement of two sheet-promoting Thr residues with Ala and elimination of an interstrand salt bridge involving the C terminus. Peptide **5a** is variant of **4a** with two destabilizing Thr→Ala substitutions.

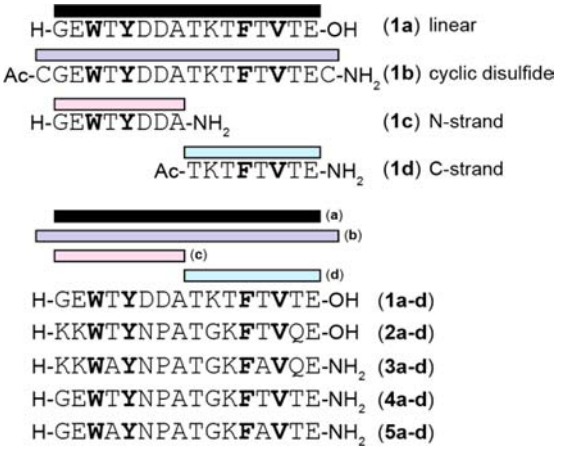


Figure 3. Hairpin model systems derived from the protein GB1. Hydrophobic core residues from the folded GB1 tertiary structure are shown in bold. For each sequence examined, four peptides were prepared to quantify the folded population by NMR: the parent sequence (**1a**–**5a**), a cyclic derivative with two added terminal Cys residues in a disulfide bridge (**1b**–**5b**), an N-terminal fragment (**1c**–**5c**), and a C-terminal fragment (**1d**–**5d**).

Prior work has established multiple techniques for the analysis of folding equilibria of short β-hairpin peptides by multidimensional NMR.[20] One commonly employed method is the comparison of a subset of H_α chemical shifts (δH_α) from a sequence of unknown folded stability to corresponding δH_α values for the same sequence in a random coil conformation; the resulting chemical shift deviations (CSDs) increase in magnitude with folded population. When combined with δH_α values for a fully folded control peptide, CSDs can be used to experimentally determine folded population.[19,21] A related method for analysis of hairpin folding by NMR involves measuring the separation of diastereotopic H_α protons in a Gly residue. In a well-folded structure, these two protons experience different chemical environments on the NMR time scale; thus, greater $H_\alpha/H_{\alpha'}$ separation is indicative of a larger folded population. Like CSD analysis, this measure can be used in concert with appropriate control sequences to quantify folded populations.[22]

In order to apply the above NMR population analyses for proposed model systems **2a**–**5a**, we prepared three variants of each peptide to be examined (Figure 3). Cyclization through intramolecular disulfide bridge involving Cys residues appended to the N and C termini provided fully folded control peptides (**2b**–**5b**).[23] Dividing the 16 residue sequences in half produced 8 residue N-terminal fragments (**2c**–**5c**) and C-terminal fragments (**2d**–**5d**) as unfolded control oligomers. We synthesized the 16 peptides in series **2**–**5** by Fmoc solid phase methods and subjected each to a panel of NMR experiments (TOCSY, COSY, NOESY) in pH 6.3 phosphate-buffered aqueous solution. Peptide **4a** showed significant signal broadening in the ^{1}H NMR (attributed to aggregation), so its folded population was not calculated. In our analysis of the data obtained for model systems **2**, **3**, and **5**, we found that H_α CSDs of residues in hydrogen-bonding positions remote from the turns were most diagnostic of folding. As noted in previous work on a related hairpin model system, chemical shift values for H_α's in non-hydrogen-bonded positions were impacted in unpredictable ways by their proximity to cross-strand aromatic residues in the folded state.[24]

We determined folded populations for peptides **2a**, **3a**, and **5a** using both H_α CSD and Gly $H_\alpha/H_{\alpha'}$ separation (Table 1).

Table 1. Population Analysis for Folding Equilibria of α-Peptide Model Systems 2a–5a

	folded population[a]	
peptide	H_α CSD[b]	Gly $H_\alpha/H_{\alpha'}$ separation[c]
2a	97 ± 6%	94 ± 6%
3a	66 ± 5%	68 ± 6%
4a	ND	ND
5a	55 ± 2%	64 ± 6%

[a]Determined by NMR analysis of a sample in 50 mM phosphate, 9:1 H_2O/D_2O, pH 6.3 at 293 K. In each case, the cyclic disulfide peptides **2b**–**5b** were used as a fully folded control. [b]Monitored by the change in H_α chemical shift relative to corresponding unfolded control sequences **2c**–**5c** and **2d**–**5d**; uncertainties represent the average population over residues 4, 11, and 13. [c]Monitored by the chemical shift separation of diastereotopic protons H_α and $H_{\alpha'}$ in Gly_{10}.

Based on these data, we ruled out model system **2** as being too stable. Sequences **3a** and **5a** showed similar folded populations, although peptide **5a** had better chemical shift dispersion and less ambiguity in resonance assignments. Model system **5** was therefore selected as the starting point for backbone modification. NMR data for the model peptides indicated that the folded populations determined by H_α CSD and glycine $H_\alpha/H_{\alpha'}$ separation were comparable within error. We employed Gly separation in subsequent population analyses to minimize the impact of changes to sequence register and hydrogen bonding resulting from different $\alpha\rightarrow\beta$ residue replacements in designs I–IV.

Peptides **1a**–**5a** share a conserved cluster of four hydrophobic residues (bold in Figure 3) that are packed in the core of the GB1 tertiary structure. Our goals in evaluating designs I–IV when applied for backbone substitution in sequence **5a** were to determine (1) whether the display of the hydrophobic core tetrad is retained after $\alpha\alpha\rightarrow\beta$ or $\alpha\alpha\rightarrow\beta\beta$ modification and (2) the thermodynamic impact of backbone modification on folding if it is tolerated. Application of design methods I–IV

dx.doi.org/10.1021/ja306311r | J. Am. Chem. Soc. 2012, 134, 15906–15913

to model hairpin **5a** generated a series of α/β-peptide analogues (**6a**–**9a**, Figure 4). Since the modifications employed

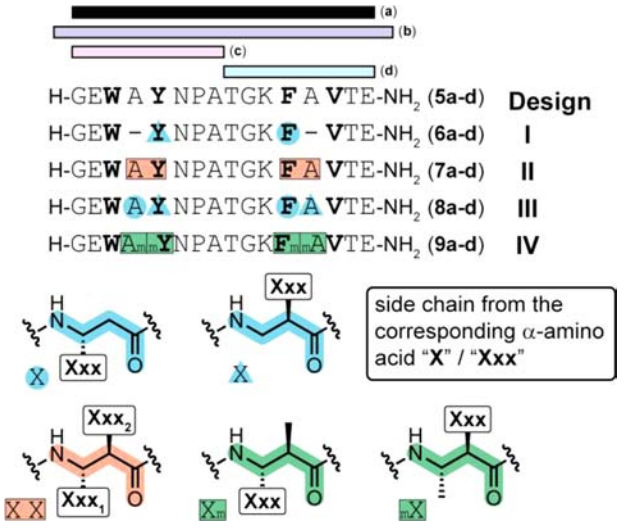


Figure 4. α/β-Peptides resulting from application of designs I–IV to α-peptide model system **5a**. For each sequence examined, four peptides were prepared to quantify the folded population by NMR: the parent sequence (**5a**–**9a**), a cyclic derivative with two added terminal Cys residues in a disulfide bridge (**5b**–**9b**), an N-terminal fragment (**5c**–**9c**), and a C-terminal fragment (**5d**–**9d**).

either extend or contract the backbone, we made matched $\alpha\alpha$ substitutions at cross-strand positions in the hairpin (residue pairs Ala_4Tyr_5 and $Phe_{12}Ala_{13}$ in **5a**). Residues Tyr_5 and Phe_{12} are found in the core of the larger GB1 protein; we therefore retained their functional groups in design I, which requires deletion of a side chain. For each backbone modification, we prepared unfolded and folded control sequences (**6b**–**9b**, **6c**–**9c**, and **6d**–**9d**), as detailed above for the α-peptide model system. Protected β-amino acids were prepared by various known routes and used in standard Fmoc solid-phase synthesis protocols (see the Supporting Information). Each peptide was purified by reverse-phase HPLC, and its identity was confirmed by mass spectrometry.

Structural Consequences of $\alpha\rightarrow\beta$ Substitution. Disulfide-cyclized peptides **6b**–**9b** were first analyzed to determine the effects of designs I–IV on the folded structure of model sequence **5b**. The Gly $H_\alpha/H_{\alpha'}$ separation for three of the α/β-peptides (**6b**, **7b**, and **9b**) was comparable to that of α-peptide **5b** (Figure 5). In contrast, peptide **8b** showed a markedly smaller separation of the diastereotopic H_α's; this observation suggests design III cannot support the hairpin fold, even with a structure-promoting disulfide bridge at the terminus. We reason that the extra flexibility imparted by the two additional methylene units resulting from $\alpha\alpha\rightarrow\beta^3\beta^2$ modification in design III destabilizes the hairpin structure.

We investigated the hydrogen-bonding pattern and side-chain display in **6b**–**9b** by analysis of long-range NOE contacts between the two strands of the hairpin. Peptides **6b** and **7b** showed a set of interstrand NOEs consistent with a nativelike hairpin fold. In support of the hypothesis underlying our design, the hydrogen-bonding pattern resulting from $\alpha\alpha\rightarrow\beta$ substitution led to a nativelike display of the hydrophobic tetrad from GB1 on one face of the sheet (Figure 6). Peptide **8b** showed no long-range NOE contacts consistent with hairpin folding, consistent with its modest Gly $H_\alpha/H_{\alpha'}$ separation.

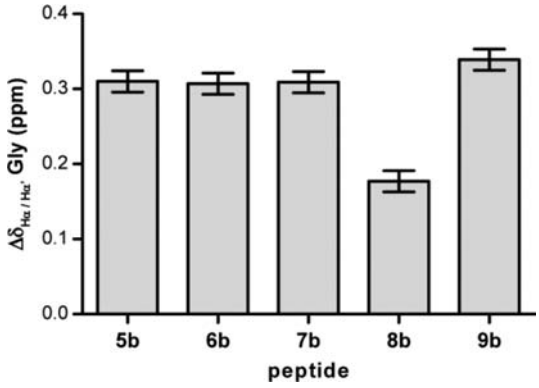


Figure 5. Glycine $H_\alpha/H_{\alpha'}$ separation for observed in the NMR of α peptide **5b** and α/β-peptide analogues **6b**–**9b**. NMR experiments were performed in 50 mM phosphate, 9:1 H_2O/D_2O, pH 6.3 at 293 K.

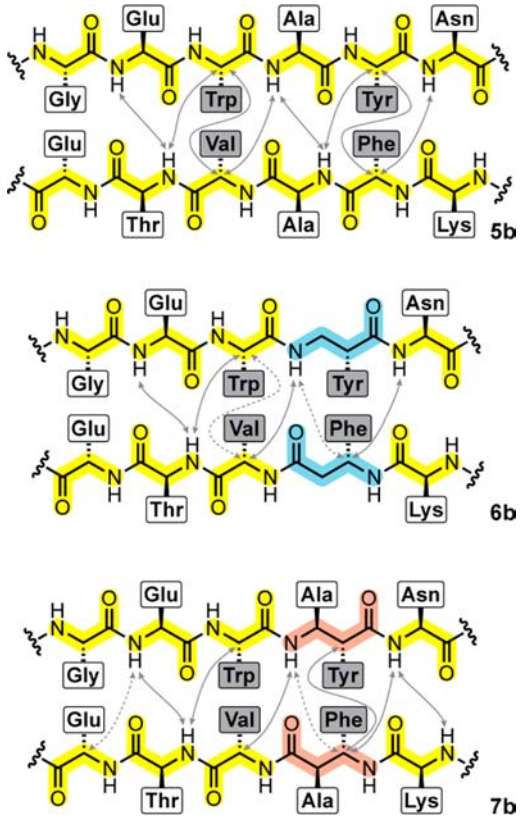


Figure 6. Key interstrand NOEs that establish hydrogen-bonding register and side-chain display in α-peptide **5b** and α/β-peptide analogues **6b** and **7b**. Dashed lines indicate NOEs with ambiguous assignments. NMR experiments were performed in 50 mM phosphate, 9:1 H_2O/D_2O, pH 6.3 at 293 K.

In contrast to the other α/β-peptides, **9b** was only sparingly soluble in the pH 6.3 buffer system employed for NMR analysis (**9b** was soluble at 0.2 mM concentration, while other NMR samples were typically prepared at 0.4–0.8 mM). The only NOE correlations observed for **9b** corresponded to very close interproton contacts (intraresidue, sequential). However, resonances were sharp and not consistent with the signal broadening typically observed with H-bonding-mediated sheet aggregation (e.g., as seen for model peptide **4a**). We performed concentration-dependent one-dimensional (1D) ^1H NMR experiments on **5b**–**9b**, and none of the oligomers showed

dx.doi.org/10.1021/ja306311r | J. Am. Chem. Soc. 2012, 134, 15906–15913

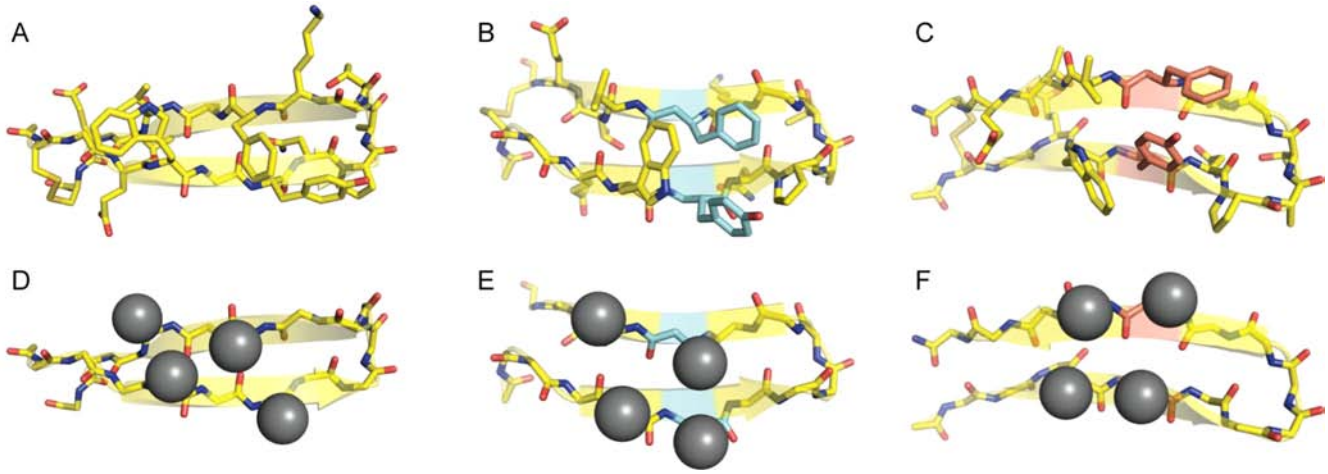


Figure 7. NMR solution structures of α-peptide **5b** and α/β-peptide analogues **6b** and **7b**. Panels (A–C) depict minimized average coordinates resulting from simulated annealing for **5b** (A), **6b** (B), and **7b** (C). Panels (D–F) compare the display of the hydrophobic tetrad from GB1 (first side-chain carbon shown as a gray sphere) in **5b** (D), **6b** (E), and **7b** (F). NMR data used for structure determination were obtained in 50 mM phosphate, 9:1 H_2O/D_2O, pH 6.3 at 293 K.

peak migration after 5× dilution of the sample used for 2D analysis. We reason that the poor aqueous solubility of **9b** may arise from a large hydrophobic patch in the folded state resulting from the insertion of four structure-promoting methyl groups in the center of the sheet. Despite its modest solubility, the glycine $H_\alpha/H_{\alpha'}$ separation data suggest that **9b** adopts a well-folded structure. We hypothesize that **9b** forms a hairpin with a nativelike display of the GB1 hydrophobic tetrad, but the lack of direct NOE evidence precludes a conclusive test of this hypothesis. Collectively, structural analysis for disulfide-cyclized peptides **6b–9b** suggests that three of our proposed designs for $\alpha\rightarrow\beta$ substitution (I, II, and IV) lead to nativelike side-chain display when applied to prototype α-peptide sheet model system **5b**.

By using NOE-derived distance restraints and simulated annealing, high-resolution NMR solution structures were determined for peptides **5a**, **5b**, **6b**, and **7b** (Figure 7 and Supporting Information, Figure S1). The calculations showed good convergence. The ensemble of low-energy structures obtained for peptides **5a**, **5b**, and **7b** each showed excellent internal agreement (Supporting Information, Figure S2). In the case of peptide **6b**, two families of structures were found: a slightly bent hairpin as well as a hairpinlike structure with a dramatic curve near the termini (Supporting Information, Figure S2). We noted some degree of twisting near the termini of all the cyclic systems, which may be necessary to accommodate the disulfide bridge. We surmise the presence of a minor folded population in **6b** with a significant kink in the backbone results from inclusion of a disulfide rather than insertion of unnatural residues. The less distorted hairpin structure of peptide **6b** was used for further analysis.

For each NMR ensemble, we calculated a minimized average structure and applied it to compare the display of the GB1 hydrophobic tetrad after backbone modification (Figure 7). Gratifyingly, the four key hydrophobic residues in α/β-peptides **6b** and **7b** are all clustered on the same face of the sheet as the natural α-peptide **5b**. This observation validates the ability of designs I and II to facilitate nativelike side-chain display in backbone modified sheets.

In order to quantitatively compare the ability of α/β-peptides **6b** and **7b** to mimic the display of the hydrophobic core residues in the natural protein, we generated overlays of each with the C-terminal hairpin from the crystal structure of full-length GB1 (Supporting Information, Figure S3).[25] Root-mean-square deviation (rmsd) values were calculated based on overlay of the backbone and first side chain C atom for each key hydrophobic residue. The rmsd values for α/β-peptides **6b** and **7b** are 1.03 and 1.80 Å, respectively. As a point of comparison, this rmsd is 1.18 Å for the corresponding overlay of α-peptide hairpin **5b** with the native protein. We attribute the better agreement of **6b** versus **7b** to the hydrophobic cluster of the native backbone to result from the higher degree of flexibility in **6b**. This increased flexibility may allow better positioning of the hydrophobic side chains after contraction of the backbone, a consequence of both designs I and II.

Thermodynamic Impact of $\alpha\rightarrow\beta$ Substitution. NMR analysis of the cyclic disulfide series **5b–9b** provided strong evidence that three of the four originally conceived designs for backbone substitution (I, II, and IV) can support hairpin folding with nativelike side-chain display when applied to a natural sequence. We next sought to compare the thermodynamic impact of the different substitution strategies on the folding equilibrium of the linear hairpin peptide **5a**. We calculated the folded population and corresponding free energy of folding ΔG_{fold} of linear peptides **5a–9a** at 278 K (Table 2) using established methods. Cyclized peptides (**5b–9b**) and strand sequences (**5c–9c**, **5d–9d**) were used as fully folded and unfolded controls, respectively. Since the cyclic disulfide

Table 2. Population Analysis for Folding Equilibria of α-Peptide 5a and α/β-Peptide Analogues 6a–9a[a]

peptide	folded population (%)	ΔG_{fold} (kcal mol^{-1})
5a	66 ± 5	−0.37 ± 0.14
6a	24 ± 5	0.65 ± 0.15
7a	17 ± 5	0.90 ± 0.19
8a	<9	>1.3
9a	22 ± 4	0.69 ± 0.14

[a]Determined by NMR analysis of a sample in 50 mM phosphate, 9:1 H_2O/D_2O, pH 6.3 at 278 K. Folded population was quantified by the chemical shift separation of diastereotopic protons H_α and $H_{\alpha'}$ in the internal Gly present in each sequence (see methods).

dx.doi.org/10.1021/ja306311r | J. Am. Chem. Soc. 2012, 134, 15906–15913

control for design III (peptide **8b**) did not appear to adopt a hairpin conformation (vide supra), we estimated a lower bound for measurable folded population of **8a** (see methods). From the above data, we tabulated the change in folding free energy ($\Delta\Delta G_{fold}$) resulting from the application of each backbone substitution strategy to α-peptide **5a** (Figure 8).

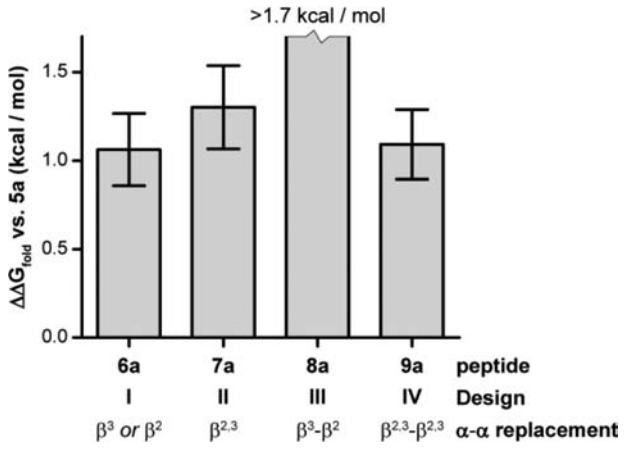


Figure 8. Thermodynamic impact of designs I–IV when applied to to α-peptide **5a**. The difference in folding free energy at 278 K relative to **5a** ($\Delta\Delta G_{fold}$) is shown for α/β-peptides **6a**–**9a**. The value for **8a** is a lower bound for the extent of destabilization.

The thermodynamic impact of designs I, II, and IV applied to parent sequence **5a** are identical, within error. Each $\alpha\alpha$ segment replacement results in a modest 0.5–0.6 kcal/mol destabilization of the folded state relative to the natural peptide. Backbone contraction ($\alpha\alpha\rightarrow\beta$ substitution) appears equally well-tolerated whether the newly introduced β-residue is conformationally flexible (β^3/β^2 in design I) or structure-promoting ($\beta^{2,3}$ in design II). We suggest that the similar folding energies of **6a** and **7a** results from two competing and offsetting factors. The possibility exists that deletion of two atoms from the native α-peptide backbone results in an unfavorable repositioning of nearby side chains. The *syn*-$\beta^{2,3}$-monomers in **7a** rigidify the backbone in an extended conformation compatible with sheet formation, but they may do so at the expense of locking in an unfavorable side-chain arrangement. The less conformationally restricted β^2- and β^3-residues in peptide **6a**, while not structure promoting, may allow for a more optimal positioning of side chains that compensates energetically for the entropic penalty resulting from enhanced backbone flexibility.

When the backbone is extended ($\alpha\alpha\rightarrow\beta\beta$ substitution), the use of structure promoting *syn*-$\beta^{2,3}$-monomers is essential to retain the hairpin fold encoded by the parent sequence. The high degree of conformational flexibility in each $\beta^3\beta^2$-segment is likely responsible for the lack of any measurable folded population for **8a**. It is also possible that extension of the backbone destabilizes the folded hairpin by disrupting diagonal hydrophobic contacts[26] between strands (e.g., Trp→Phe interactions seen in **5b** and **6b**). The structure-promoting monomers in **9a** mitigate the energetic penalty resulting from backbone conformational freedom but still prevent the diagonal hydrophobic contacts that may stabilize the hairpin.

CONCLUSIONS

In summary, we have described here four different design strategies for the sequence-based backbone modification of protein β-sheets. We applied each of these designs to a model α-peptide hairpin derived from the sheet domain of a protein tertiary fold. NMR structural analysis shows that three of the four designs, which are based on $\alpha\rightarrow\beta$ residue replacement, lead to α/β-peptides that adopt nativelike hairpin folds in aqueous buffer. Importantly and in contrast to prior work,[7] each modified backbone displays a key hydrophobic cluster from the native sequence on one face of the unnatural sheet. Thermodynamic analyses of folding equilibria show that the α/β-peptide hairpins are destabilized by only 0.5–0.6 kcal/mol per $\alpha\alpha$-segment substitution.

Based on the findings described here, we propose that designs I, II, and IV may be tolerated in a larger protein prototype that includes β-sheet secondary structures. The small free-energy penalties to folding, significant in the context of a marginally stable hairpin peptide, might be more easily tolerated in a well-folded protein, where there is more free energy to spare. However, it is important to note the gap in complexity between a hairpin and a sheet in a tertiary context. The development of foldamers with structural complexity of typical protein tertiary folds is a significant and unmet challenge in the field. The results we report here do not achieve this goal, but they represent an important advance toward it. Taken along with precedent on the sequence-based mimicry of helices,[4−6] turns,[10] and loops,[11] the new insights into sheet mimicry reported here lay the groundwork for the design of tertiary structure mimics based on readily available amino acid sequence information. We are currently working to develop and test such a design strategy. These studies will provide a crucial test of the design principles developed here for sheet mimicry by establishing their applicability in the much more demanding context of a folded protein tertiary structure. Moreover, they will take us one step closer to the long-term goal of a general methodology for the design of unnatural analogues of biologically active proteins with improved activity and/or biostability.

ASSOCIATED CONTENT

Supporting Information

Figures S1–S3, Tables S1–S17, experimental methods, and minimized average coordinates for the NMR solution structures of **5a**, **5b**, **6b**, and **7b**. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

horne@pitt.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Funding for this work was provided by the University of Pittsburgh.

REFERENCES

(1) Gellman, S. H. *Acc. Chem. Res.* **1998**, *31*, 173–180.

(2) (a) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. *Chem. Rev.* **2001**, *101*, 3893–4012. (b) Davis, J. M.; Tsou, L. K.; Hamilton, A. D. *Chem. Soc. Rev.* **2007**, *36*, 326–334. (c) Goodman, C. M.; Choi, S.; Shandler, S.; DeGrado, W. F. *Nat. Chem. Biol.* **2007**, *3*, 252–262. (d) Bautista, A. D.; Craig, C. J.; Harker, E. A.; Schepartz, A. *Curr. Opin. Chem. Biol.* **2007**, *11*, 685–692. (e) Guichard, G.; Huc, I.

Chem. Commun. **2011**, *47*, 5933−5941. (f) Horne, W. S. *Expert Opin. Drug Discovery* **2011**, *6*, 1247−1262.
(3) (a) Horne, W. S.; Gellman, S. H. *Acc. Chem. Res.* **2008**, *41*, 1399−1408. (b) Pilsl, L. K. A.; Reiser, O. *Amino Acids* **2011**, *41*, 709−718.
(4) (a) Horne, W. S.; Price, J. L.; Keck, J. L.; Gellman, S. H. *J. Am. Chem. Soc.* **2007**, *129*, 4178−4180. (b) Horne, W. S.; Price, J. L.; Gellman, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9151−9156. (c) Giuliano, M. W.; Horne, W. S.; Gellman, S. H. *J. Am. Chem. Soc.* **2009**, *131*, 9860−9861. (d) Price, J. L.; Horne, W. S.; Gellman, S. H. *J. Am. Chem. Soc.* **2010**, *132*, 12378−12387.
(5) (a) Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellman, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 14751−14756. (b) Johnson, L. M.; Horne, W. S.; Gellman, S. H. *J. Am. Chem. Soc.* **2011**, *133*, 10038−10041. (c) Johnson, L. M.; Mortenson, D. E.; Yun, H. G.; Horne, W. S.; Ketas, T. J.; Lu, M.; Moore, J. P.; Gellman, S. H. *J. Am. Chem. Soc.* **2012**, *134*, 7317−7320.
(6) (a) Horne, W. S.; Boersma, M. D.; Windsor, M. A.; Gellman, S. H. *Angew. Chem., Int. Ed.* **2008**, *47*, 2853−2856. (b) Lee, E. F.; Smith, B. J.; Horne, W. S.; Mayer, K. N.; Evangelista, M.; Colman, P. M.; Gellman, S. H.; Fairlie, W. D. *ChemBioChem* **2011**, *12*, 2025−2032. (c) Boersma, M. D.; Haase, H. S.; Peterson-Kaufman, K. J.; Lee, E. F.; Clarke, O. B.; Colman, P. M.; Smith, B. J.; Horne, W. S.; Fairlie, W. D.; Gellman, S. H. *J. Am. Chem. Soc.* **2012**, *134*, 315−323.
(7) Lengyel, G. A.; Frank, R. C.; Horne, W. S. *J. Am. Chem. Soc.* **2011**, *133*, 4246−4249.
(8) Haase, H. S.; Peterson-Kaufman, K. J.; Lan Levengood, S. K.; Checco, J. W.; Murphy, W. L.; Gellman, S. H. *J. Am. Chem. Soc.* **2012**, *134*, 7652−7655.
(9) (a) David, R.; Gunther, R.; Baumann, L.; Luhmann, T.; Seebach, D.; Hofmann, H.-J.; Beck-Sickinger, A. G. *J. Am. Chem. Soc.* **2008**, *130*, 15311−15317. (b) Lee, B.-C.; Zuckermann, R. N. *ACS Chem. Biol.* **2011**, *6*, 1367−1374.
(10) (a) Baca, M.; Kent, S. B. H.; Alewood, P. F. *Protein Sci.* **1993**, *2*, 1085−1091. (b) Viles, J. H.; Patel, S. U.; Mitchell, J. B. O.; Moody, C. M.; Justice, D. E.; Uppenbrink, J.; Doyle, P. M.; Harris, C. J.; Sadler, P. J.; Thornton, J. M. *J. Mol. Biol.* **1998**, *279*, 973−986. (c) Odaert, B.; Jean, F.; Melnyk, O.; Tartar, A.; Lippens, G.; Boutillon, C.; Buisine, E. *Protein Sci.* **1999**, *8*, 2773−2783. (d) Kaul, R.; Angeles, A. R.; Jäger, M.; Powers, E. T.; Kelly, J. W. *J. Am. Chem. Soc.* **2001**, *123*, 5206−5212. (e) Arnold, U.; Hinderaker, M. P.; Nilsson, B. L.; Huck, B. R.; Gellman, S. H.; Raines, R. T. *J. Am. Chem. Soc.* **2002**, *124*, 8522−8523. (f) Kaul, R.; Deechongkit, S.; Kelly, J. W. *J. Am. Chem. Soc.* **2002**, *124*, 11900−11907. (g) Tam, A.; Arnold, U.; Soellner, M. B.; Raines, R. T. *J. Am. Chem. Soc.* **2007**, *129*, 12670−12671. (h) Liu, F.; Du, D.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2369−2374. (i) Fuller, A. A.; Du, D.; Liu, F.; Davoren, J. E.; Bhabha, G.; Kroon, G.; Case, D. A.; Dyson, H. J.; Powers, E. T.; Wipf, P.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11067−11072.
(11) (a) Green, B. R.; Catlin, P.; Zhang, M. M.; Fiedler, B.; Bayudan, W.; Morrison, A.; Norton, R. S.; Smith, B. J.; Yoshikami, D.; Olivera, B. M.; Bulaj, G. *Chem. Biol.* **2007**, *14*, 399−407. (b) Valverde, I. E.; Lecaille, F.; Lalmanach, G.; Aucagne, V.; Delmas, A. F. *Angew. Chem., Int. Ed.* **2012**, *51*, 718−722. (c) Reinert, Z. E.; Musselman, E. D.; Elcock, A. H.; Horne, W. S. *ChemBioChem* **2012**, *13*, 1107−1111.
(12) (a) Krauthauser, S.; Christianson, L. A.; Powell, D. R.; Gellman, S. H. *J. Am. Chem. Soc.* **1997**, *119*, 11719−11720. (b) Seebach, D.; Abele, S.; Gademann, K.; Jaun, B. *Angew. Chem., Int. Ed.* **1999**, *38*, 1595−1597. (c) Karle, I. L.; Gopi, H. N.; Balaram, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3716−3719. (d) Karle, I.; Gopi, H. N.; Balaram, P. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5160−5164. (e) Martinek, T. A.; Toth, G. K.; Vass, E.; Hollosi, M.; Fulop, F. *Angew. Chem., Int. Ed.* **2002**, *41*, 1718−1721.
(13) (a) Chi, Y.; Gellman, S. H. *J. Am. Chem. Soc.* **2006**, *128*, 6804−6805. (b) Chi, Y.; English, E. P.; Pomerantz, W. C.; Horne, W. S.; Joyce, L. A.; Alexander, L. R.; Fleming, W. S.; Hopkins, E. A.; Gellman, S. H. *J. Am. Chem. Soc.* **2007**, *129*, 6050−6055.
(14) Xu, X.; Wang, K.; Nelson, S. G. *J. Am. Chem. Soc.* **2007**, *129*, 11690−11691.
(15) Langenhan, J. M.; Gellman, S. H. *J. Org. Chem.* **2003**, *68*, 6440−6443.
(16) Wüthrich, K. *NMR in Structural Biology: A Collection of Papers by Kurt Wüthrich*; World Scientific: River Edge, NJ, 1995.
(17) (a) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **1998**, *54*, 905−921. (b) Brunger, A. T. *Nat. Protocols* **2007**, *2*, 2728−2733.
(18) (a) Searle, M. S.; Ciani, B. *Curr. Opin. Struct. Biol.* **2004**, *14*, 458−464. (b) Nowick, J. S. *Acc. Chem. Res.* **2008**, *41*, 1319−1330.
(19) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Mol. Biol.* **1994**, *1*, 584−590.
(20) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.* **2004**, *126*, 7238−7243.
(21) Fesinmeyer, R.; Hudson, F.; Olsen, K.; White, G.; Euser, A.; Andersen, N. *J. Biomol. NMR* **2005**, *33*, 213−231.
(22) Griffiths-Jones, S. R.; Maynard, A. J.; Searle, M. S. *J. Mol. Biol.* **1999**, *292*, 1051−1069.
(23) Tatko, C. D.; Waters, M. L. *Protein Sci.* **2003**, *12*, 2443−2452.
(24) Espinosa, J. F.; Gellman, S. H. *Angew. Chem., Int. Ed.* **2000**, *39*, 2330−2333.
(25) Frericks Schmidt, H. L.; Sperling, L. J.; Gao, Y. G.; Wylie, B. J.; Boettcher, J. M.; Wilson, S. R.; Rienstra, C. M. *J. Phys. Chem. B* **2007**, *111*, 14362−14369.
(26) Syud, F. A.; Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **2001**, *123*, 8667−8677.